Amendments to the Specification:

On page 6, the 2nd paragraph, beginning on line 9, is amended as follows:

In an embodiment of the present invention, the data warehouse may be modeled as separate sample, gene annotation, and gene expression multi-dimensional data spaces. Basic operations in these data spaces in terms of traditional on-line analytical processing ("OLAP") dimension reduction and aggregation manipulations may be used for complex gene expression analysis operations.. Data warehouse management tools are used for maintaining data consistency, with process specific consistency rules checking the correct execution of data migration and integration processes and with domain specific rules validating sample, expression, and gene annotation data. In accordance with one embodiment of the present invention, an archive may be used to provide a uniform analysis interface for gene expression data from alternate gene expression databases, such as the Genbank public domain database available on the ~~Internet at www.ncbi.nlm.nih.gov/Genbank~~ World Wide Web at ncbi.nlm.nih.gov/Genbank

On page 14, the 4th paragraph, beginning on line 21, is amended as follows:

In a metabolic pathway, the components represent enzymatic activities that may be identified by EC numbers. Strongly and weakly expressed genes encoding enzymes are darkly and lightly shaded, respectively. Multiple genes may code for enzymes with the same activity, such as the many different alcohol dehydrogenases. In addition, multiple fragments may represent the same gene. The underlying pathway diagrams may be obtained from a public source, such as KEGG available ~~at www.genome.ed.jp/kegg~~ on the World Wide Web at genome.ed.jp/kegg. Pathway visualizations may be performed for a particular sample set and gene set. The gene set may be computed indirectly from sample sets using the Gene Signature tool, Gene Signature Differential or Fold Change Analysis tools, or may be selected directly.

On page 19, the 3rd paragraph, beginning on line 13, is amended as follows:

As those skilled in the art should appreciate, GenBank is the National Institutes of Health ("NIH") genetic sequence database, an annotated collection of all publicly available DNA sequences that is available on the ~~Internet at www.ncbi.nlm.nih.gov/Genbank~~ World Wide Web at ncbi.nlm.nih.gov/Genbank. In addition, UniGene is a system for automatically partitioning GenBank sequences into a non-redundant set of gene-oriented clusters and is available ~~at www.ncbi.nlm.nih.govfUniGene/~~ on the World Wide Web at ncbi.nlm.nih.gov/UniGene. Finally, LocusLink provides a single query interface to curated sequence and descriptive information about genetic loci and is available ~~at www.locuslink.com~~ on the World Wide Web at locuslink.com. LocusLink presents information on official nomenclature, aliases, sequence accessions, phenotypes, EC numbers, MIM numbers, UniGene clusters, homology, map locations, and related web sites.

On page 41, the 1st paragraph, beginning on line 2, is amended as follows:

In another preferred embodiment of the present invention, the staging database is a proper relational database with SQL query capability. The staging database preferably also provides reports to track the staging activity. Such reports include a staging loading ~~eport,~~ report that is issued any time loading to the staging database occurs; a staging weekly report ~~which~~ that reports the staging activity per week, i.e., number of experiments loaded in, number of experiments migrated to the relational database, etc.; and a staging weekly exception report ~~which~~ that reviews double scan experiments[[,]] and reports the experiment names of experiments waiting for the "mate" scan (are on hold) for longer than 5 days.

On page 43, the 1st paragraph, beginning on line 4, is amended as follows:

Chip consistency rules assess the microarray for consistency and are preferably checked at the time of publishing and data staging. Chip defects report consistency rules assess the chip defects report for consistency. For example, the gene fragment names in the chip defects report per experiment should match the gene fragment names of the chip type in the experiment. Clinical data consistency rules assess the internal consistency of the clinical data. Clinical data/gene expression data consistency assess the consistency of the clinical data with the gene

expression data. For example, the organ name in the clinical database should match the target type value in the gene expression data for the same sample. Matching is preferably performed at variable granularity, i.e., organ "cerebellum" matches target type "brain". Fragment/gene expression data consistency assesses the consistency of the fragment index data with the gene expression data. Preferably, this rule verifies that the ID and ITEM_NAME in BIOLOGICAL_ITEM joined with the ANALYSIS_SCHEME.ID, matches the ITEM_ID, AFFY_NAME and ON_CHIP attributes of the fragment index's AFFY_NAME. Expression integrity rules are based on biological knowledge. For example, if a gene is known to be present in a specific tissue type, then it should be present in the relational database. Special classes of ~~this~~ these rules handle the housekeeping (or spiking) genes for which there is prior knowledge as ~~of~~ to whether they are present or absent. Figure 8 represents an embodiment of the integrity constraint enforcement system of the present invention. The application-specific rules and general biological rules are organized by modules, 801 and 802, and are stored in the Rule Repository 800. When an application-specific or general biological function is run and an error is detected, then the system generates an error codes and/or corrects the error by means of the error engine 803. In addition, a log and audit engine 804 creates a log and audit of the run.